A Randomized, Placebo-Controlled, Double-Blind, Multicenter Study to Evaluate Efficacy and Safety of Oral BT-11 in Mild to Moderate Ulcerative Colitis

BT-11-201, Phase 2

STATISTICAL ANALYSIS PLAN

Version 01
November 3, 2020

Status: Effective

Prepared by: Matthew Rueffer
Alimentiv Inc.
100 Dundas St. Suite 200
London Ontario
Canada
N6A 5B6

Sponsor: Landos Biopharma Inc.
1800 Kraft Drive, Suite 216
Blacksburg, VA,
USA
24060

APPROVAL

A Randomized, Placebo-Controlled, Double-Blind, Multicenter Study to Evaluate Efficacy and Safety of Oral BT-11 in Mild to Moderate Ulcerative Colitis

BT-11-201, Phase 2

STATISTICAL ANALYSIS PLAN

Version 01
November 3, 2020

Status: Effective

Name	Position	Signature	Date
[1] Matthew Rueffer	Manager, Statistics and SAS Programming, Alimentiv Inc.	Matthew Rueffer	03 NOV 2020
[2] Kristina Mardinian	Statistician Alimentiv Inc.	Kristina Mardinian	04 NOV 2020
[3] Josep Bassaganya-Riera	Study Director, Landos Biopharma Inc.	DocuSigned by: Josep Bassaganya-Riera Signer Name: Josep Bassaganya-Riera Signing Reason: I approve this document Signing Time: 11/3/2020 | 7:51:22 PM PST 04CC2C7580B1482B93401F38FA13DA1C	

[1] Author, signs for correctness and completeness

[2] Reviewer, signs for correctness and completeness

[3] Approver, signs for the release of the document

TABLE OF CONTENTS

DEFINITIONS / ABBREVIATIONS

ADR	adverse drug reaction
AE	adverse event
ALT	alanine aminotransferase
AST	aspartate aminotransferase
BMI	body mass index
ECG	electrocardiogram
eCRF	electronic case report form
EOT	end of treatment
FOXP3	forkhead box P3
hs-CRP	high-sensitivity CRP
IBD	inflammatory bowel disease
IBDQ	Inflammatory Bowel Disease Questionnaire
IFNγ	interferon-γ
IL	interleukin
INR	international normalized ratio
IWRS	interactive web-based response system
LANCL2	lanthionine synthetase C-like 2
LPMC	lamina propria mononuclear cell
MCP1	monocyte chemoattractant protein-1
MCS	Mayo Clinic Score
MedDRA	Medical Dictionary for Regulatory Activities
MES	Mayo endoscopic subscore
mITT	modified intent-to-treat
NK	natural killer (cells)
OLE	open-label extension
PBMC	peripheral blood mononuclear cell
PK	pharmacokinetic
PP	per-protocol
qPCR	quantitative polymerase chain reaction
RHI	Robarts Histopathology Index
SAE	serious adverse event
SAP	statistical analysis plan
TBL	total bilirubin level
Th	T helper (cells)

TNF	tumor necrosis factor
Treg	T regulatory (cells)
UC	ulcerative colitis
ULN	upper limit of normal
US	United States

1 INTRODUCTION

The purpose of this statistical analysis plan is to describe in detail the analyses to be performed for the induction, maintenance, and open-label extension phases of the BT-11-201 study. This plan does not cover the analysis of the biomarker endpoints. The approval of this statistical analysis plan must be obtained prior to the unblinding of treatment allocation.

2 SUMMARY OF STUDY PROTOCOL

2.1 STUDY OBJECTIVES

2.1.1 Primary Objective

The primary objective of this study is to establish efficacy and safety of oral BT-11 in inducing clinical remission at Week 12 in subjects with mild to moderate ulcerative colitis (UC).

2.1.2 Secondary Objectives

The secondary objectives of this study are to evaluate the following at Week 12:

- To evaluate the effects of BT-11 on disease activity as measured by symptoms, endoscopy, histology, and biomarkers during the 12-week induction period
- To assess health-related quality of life during the 12-week induction period
- To assess the PK parameters of BT-11 during the 12-week induction period
- To evaluate safety during the 12-week induction period

2.1.3 Exploratory Objectives

The following exploratory objectives relate to the evaluation of endpoints during the maintenance phase:

- To evaluate the effects of BT 11 on disease activity as measured by symptoms, endoscopy, histology, and biomarkers during up to 30 weeks of maintenance therapy
- To assess health-related quality of life during up to 30 weeks of maintenance therapy
- To assess the PK parameters of BT-11 during up to 30 weeks of maintenance therapy
- To evaluate safety during up to 30 weeks of maintenance therapy
- To evaluate target engagement and mechanism of action
- To explore association of drug exposure in colonic mucosal tissue biopsies with clinical, endoscopic, histopathologic, cellular, and molecular

The following exploratory objectives relate to the evaluation of endpoints during the open-label extension phase:

- To assess the safety of the extended use of oral BT-11 in subjects with mild to moderate ulcerative colitis
- To evaluate the effects of BT 11 on measures of efficacy

2.2 OUTCOME MEASURES

The following table provides a summary of the outcome measurements for this study.

2.2.1 Induction Outcomes

	Objective	Endpoint	Summary Measures
Primary	To assess the efficacy and safety of oral BT-11 in inducing clinical remission at Week 12 in subjects with mild to moderate UC	• Clinical remission rate at Week 12, defined using the 3-component modified Mayo Score as a rectal bleeding subscore of 0, a stool frequency subscore of 0 or 1, and an endoscopic subscore of 0 or 1	Difference in proportions between 1,000 mg dose group vs. placebo group Followed sequentially by difference in proportions between 500 mg dose group vs. placebo group
Secondary	To evaluate the effects of BT-11 on disease activity as measured by symptoms, endoscopy, histology, and biomarkers during the 12-week induction period	Key (ranked) secondary endpoints: • Endoscopic remission rate at Week 12, defined as a Mayo endoscopic subscore [MES] of 0 or 1 • Endoscopic response rate at Week 12, defined as a decrease from baseline in MES of ≥ 1 point • Clinical response rate at Week 12, defined as decrease from baseline in Mayo Score of ≥ 3 points and ≥ 30 percent, with an accompanying decrease in the subscore for rectal bleeding of ≥ 1 point or an absolute subscore for rectal bleeding of 0 or 1 • Mucosal healing rate at Week 12, defined by a MES of 0 or 1 and a Geboes score < 3.1 • Histologic remission rate at Week 12, defined by a Geboes score < 3.1	Difference in proportions between 1,000 mg dose group vs. placebo group Followed sequentially by difference in proportions between 500 mg dose group vs. placebo group

	Objective	Endpoint	Summary Measures
Secondary	To evaluate the effects of BT-11 on disease activity as measured by symptoms, endoscopy, histology, and biomarkers during the 12-week induction period	Other secondary endpoints: • Mean change in 3-component modified Mayo score from baseline to Week 12 • Mean change in partial Mayo Score from baseline to Weeks 2, 6, and 12 • Mean change in Mayo rectal bleeding subscore from baseline to Weeks 2, 6, and 12 • Mean change in Mayo stool frequency subscore from baseline to Weeks 2, 6, and 12 • Mean change in MES from baseline to Week 12 • Mean change in Robarts Histopathology Index (RHI) scores from baseline to Week 12 • Mean change in fecal calprotectin from baseline to Weeks 2, 6, and 12	Difference in means between 1,000 mg and 500mg dose groups (separately) vs. placebo group
		• Clinical remission rate at Week 12, based on alternate definition ("MCS clinical remission") of total Mayo Score ≤ 2 with all subscores ≤ 1 • Normalization of fecal calprotectin at Weeks 2, 6, and 12 in subjects with abnormal fecal calprotectin at baseline (abnormal defined as fecal calprotectin > 250 mg/kg) • Normalization of hs-CRP at Weeks 2, 6, and 12 in subjects with abnormal hs-CRP at baseline (abnormal defined as hs-CRP > 3.0 mg/L)	Difference in proportions between 1,000 mg and 500 mg dose groups (separately) vs. placebo group
		• Mean change in high-sensitivity C-reactive protein (hs-CRP) from baseline to Weeks 2, 6, and 12 • Mean change in UC-100 score from baseline to Week 12 • Change in Robarts Symptom and Impacts Questionnaire for Ulcerative Colitis (SIQ-UC) items from baseline to Week 12	Difference in means between 1,000 mg and 500 mg dose groups (separately) vs. placebo group

	Objective	Endpoint	Summary Measures
Secondary	To assess health-related quality of life during the 12-week induction period	• Mean change in Inflammatory Bowel Disease Questionnaire (IBDQ) score from baseline to Week 12	Difference in means between 1,000 mg and 500 mg dose groups (separately) vs. placebo group
	To assess the PK parameters of BT-11 during the 12-week induction period	• In the active BT-11 treatment group, BT-11 concentration in serum, feces, and tissue	Difference in means between 1,000 mg and 500 mg dose groups (separately) vs. placebo group
	To evaluate safety during the 12-week induction period	• Frequency and severity of AEs	Difference in proportions and relative risk between 1,000 mg and 500 mg dose groups (separately) vs. placebo group
		• Changes in clinical chemistry and hematology from baseline • Results of vital signs and physical examination • ECG findings	Difference in means between 1,000 mg and 500 mg dose groups (separately) vs. placebo group

Abbreviations: AE, adverse event; ECG, electrocardiogram; hs-CRP, high-sensitivity C-reactive protein; IBDQ, Inflammatory Bowel Disease Questionnaire; MES, Mayo endoscopic subscore; RHI, Robarts Histopathology Index.

2.2.2 Maintenance Phase Outcomes

	Objectives	Endpoints	Summary Measures
Exploratory	To evaluate the effects of BT 11 on disease activity as measured by symptoms, endoscopy, histology, and biomarkers during up to 30 weeks of maintenance therapy	• Durable clinical remission, defined as clinical remission at both Week 12 and Week 30 • Durable clinical remission (defined as MCS clinical remission at both Week 12 and Week 30) • Durable clinical response, defined as clinical response at both Week 12 and Week 30 • Endoscopic remission rate at Week 30, defined by a MES of 0 or 1 • Endoscopic response rate at Week 30, defined as a decrease from baseline in MES of ≥ 1 point • Corticosteroid-free clinical remission at Week 30 • Corticosteroid-free endoscopic remission at Week 30 • Mucosal healing rate at Week 30, defined by a MES of 0 or 1 and a Geboes score < 3.1 • Histologic remission rate at Week 30, defined by a Geboes score < 3.1 • Clinical response rate at Week 30, defined as a decrease from baseline in Mayo Score of ≥ 3 points and ≥ 30 percent, with an accompanying decrease in the subscore for rectal bleeding of ≥ 1 point or an absolute subscore for rectal bleeding of 0 or 1	Difference in proportions between 1,000 mg and 500 mg dose groups (separately) vs. placebo group

	Objectives	Endpoints	Summary Measures
Exploratory	To evaluate the effects of BT 11 on disease activity as measured by symptoms, endoscopy, histology, and biomarkers during up to 30 weeks of maintenance therapy	• Mean change in 3-component modified Mayo score from baseline to Week 30. • Mean change in partial Mayo Score from baseline to Weeks 18, 24, and 30 • Mean change in Mayo rectal bleeding subscore from baseline to Weeks 18, 24, and 30 • Mean change in Mayo stool frequency subscore from baseline to Weeks 18, 24, and 30 • Mean change in MES from baseline to Week 30 • Mean change in RHI scores from baseline to Week 30 • Mean change in fecal calprotectin from baseline to Weeks 18, 24, and 30	Difference in means between 1,000 mg and 500 mg dose groups (separately) vs. placebo group
		• Normalization of fecal calprotectin at Weeks 18, 24, and 30 in subjects with abnormal fecal calprotectin at baseline (abnormal defined as fecal calprotectin > 250 mg/kg) • Normalization of hs-CRP at Weeks 18, 24, and 30 in subjects with abnormal hs-CRP at baseline (abnormal defined as hs-CRP > 3.0 mg/L)	Difference in proportions between 1,000 mg and 500 mg dose groups (separately) vs. placebo group
		• Mean change in hs-CRP from baseline to Weeks 18, 24, and 30 • Mean change in UC-100 score from baseline to Week 30 • Change in Robarts Symptom and Impacts Questionnaire for Ulcerative Colitis (SIQ-UC) items from baseline to Week 30	Difference in means between 1,000 mg and 500 mg dose groups (separately) vs. placebo group
	To assess health-related quality of life during up to 30 weeks of maintenance therapy	• Mean change in IBDQ score from baseline to Week 30	Difference in means between 1,000 mg and 500 mg dose groups (separately) vs. placebo group

	Objectives	Endpoints	Summary Measures
Exploratory	To assess the PK parameters of BT-11 during up to 30 weeks of maintenance therapy	• In the active BT-11 treatment group, BT-11 concentration in serum, feces, and tissue	Difference in means between 1,000 mg and 500 mg dose groups (separately) vs. placebo group
	To evaluate safety during up to 30 weeks of maintenance therapy	• Frequency and severity of AEs	Difference in means between 1,000 mg and 500 mg dose groups (separately) vs. placebo group
		• Changes in clinical chemistry and hematology from baseline • Results of vital signs and physical examination • ECG findings	Difference in means between 1,000 mg and 500 mg dose groups (separately) vs. placebo group
	To evaluate target engagement and mechanism of action†	• Immunohistochemistry on formalin-fixed, paraffin-embedded colonic mucosal tissue biopsies • MPO (marker for neutrophil infiltration in tissue) • CD25 (Treg cell marker) • FOXP3 (Treg cell marker) • LANCL2 (target) • Multiplex cytokine analysis in serum and tissue - TNFα, IFNγ, IL-4, MCP1, MIP1α, IL-8, IL-6 (pro-inflammatory cytokines) - IL-10 (anti-inflammatory cytokine) • Transcriptomic analysis on colonic mucosal tissue biopsies using quantitative polymerase chain reaction (qPCR) • TNFα, IFNγ, IL-4, MCP1, IL-8, IL-6 (pro-inflammatory cytokines) • IL-10, FOXP3 (anti-inflammatory cytokines) • LANCL2 (target) • Flow cytometry on colonic mucosal tissue biopsies • Th1 cells • Neutrophils • IL-10-producing cellular subsets	Summary of mean for observed and change from baseline values

		including CX3CR1+ macrophages • Treg cells	
	Objectives	**Endpoints**	**Summary Measures**
Exploratory	To explore association of drug exposure in colonic mucosal tissue biopsies with clinical, endoscopic, histopathologic, cellular, and molecular outcomes†	• Exposure-response analysis	Correlation analysis

Abbreviations: CX3CR1, CX3C chemokine receptor 1; FOXP3, forkhead box P3 (protein); hs-CRP, high-sensitivity C-reactive protein; IL, interleukin; IFNγ, interferon gamma; LANCL2, lanthionine synthetase C-like 2; MCP1, monocyte chemoattractant protein-1; MIP1α, macrophage inflammatory protein-1 α; MPO, myeloperoxidase; PK, pharmacokinetic; TNFα, tumor necrosis factor alpha; Treg, T regulatory cells.

†Details of these analyses are not covered by this plan.

2.2.3 Open-Label Phase Outcomes

	Objectives	**Endpoints**	**Summary Measures**
Exploratory	To assess the safety of the extended use of oral BT-11 in subjects with mild to moderate ulcerative colitis	• Percentage of subjects with treatment-emergent adverse events (TEAEs), serious adverse events (SAEs), and TEAEs that lead to discontinuation of study drug at each study visit	Summary of proportions
		• Results of vital signs and physical examination	Summary of mean for observed and change from baseline values

	Objectives	Endpoints	Summary Measures
	To evaluate the effects of BT-11 on measures of efficacy (based on subgroups defined in the statistical analysis plan)	• Partial Mayo Score clinical remission, defined as partial Mayo Score <2, at each study visit • Partial Mayo Score clinical response, defined as partial Mayo Score decrease of ≥ 2 points from Day 1, at each study visit • Mean change in partial Mayo Score from Day 1 to each study visit • Mean change in Mayo rectal bleeding subscore from Day 1 to each study visit • Mean change in Mayo stool frequency subscore from Day 1 to each study visit	Summary of mean for observed and change from baseline values

2.3 STUDY DESIGN

This is a phase 2 randomized, placebo-controlled, double-blind, parallel-group, multicenter study. The purpose of this study is to evaluate the efficacy and safety of oral BT-11 compared to placebo in subjects with mild to moderate UC. Approximately 46 sites will participate from Europe and the USA.

A total of 195 subjects with mild to moderate UC (total Mayo Score 4-10; Mayo endoscopic subscore [MES] ≥ 2) will be randomized in a 1:1:1 ratio to receive BT-11 low-dose (500 mg), BT-11 high-dose (1,000 mg) or placebo. Each of the treatment arms will comprise 65 subjects. The randomization will be stratified by prior exposure to biologic therapy for UC (yes/no) and corticosteroid use at baseline (yes/no).

The study consists of a 28-day screening period, a 12-week induction period, an 18-week maintenance extension period, and a 2-week post-treatment safety follow-up period.

The analysis of induction of clinical remission (induction data lock and analysis) will be conducted after all subjects have reached Week 12. Eligible subjects who meet the definition of clinical response and/or clinical remission at Week 12 will be classified as responders and will continue on to be followed in the maintenance period of the trial. Subjects who are non- responders at Week 12 may be eligible for a possible separate open-label extension (OLE) study.

Subjects will be randomized to receive BT-11 low-dose (500 mg), BT-11 high-dose (1,000 mg), or placebo once-daily for 12 weeks during the induction period of the study. Subjects who continue on to the maintenance period will remain in the same blinded treatment group to which they were originally randomized.

Authorized personnel at the investigative site will administer the first dose of the study drug or placebo in

a blinded fashion. All tablets administered (placebo and BT-11) will have the same appearance and size. Each subject will receive blister packs of the study drug (high-dose BT-11, low-dose BT-11, or placebo).

2.4 PROTOCOL VIOLATIONS

Review of all major and minor protocol violations will be performed on an ongoing basis during the conduct of the study.

Major protocol violations will be excluded from the Per Protocol (PP) population. All protocol exemptions and violations that require medical review will be listed. Study subjects will be assigned to the appropriate analysis population(s) prior to the declaration of final database. The records of protocol violations will be merged and incorporated into the final database prior to database lock.

2.5 INCLUSION/EXCLUSION CRITERIA

Subjects who did not meet the following inclusion/exclusion criteria are considered major protocol violations and will be excluded from the PP population. Key inclusion criteria: male and female subjects aged 18 to 75 years with a diagnosis of UC for at least 3 months; mild to moderate UC defined by a total Mayo Score of 4 to 10 with MES ≥ 2 (confirmed by central reader); prior biologic must have stopped at least 8 weeks before study and previous biologic treatment failure is limited to one class of biologic; 5-aminosalicylates (max 4.8 g/day) and oral corticosteroids (max 20mg/day prednisone or equivalent) must be stable for the duration of the trial.

Key exclusion criteria: severe UC defined by modified Truelove and Witts criteria; disease activity limited to distal 15cm (proctitis); treatment with immunosuppressant within 25 days prior to randomization; current bacterial or parasitic pathogenic enteric infection; live virus vaccination within 1 month prior to screening.

For the maintenance period, key inclusion criteria include meeting eligibility requirements for clinical response at Week 12 (defined as decrease from baseline in Mayo Score of at least 3 points and at least 30 percent, with an accompanying decrease in the subscore for rectal bleeding of at least 1 point or an absolute subscore for rectal bleeding of 0 or 1), the ability to begin the maintenance period within 2 weeks of completing the induction Week 12 visit, and maintenance of stable doses of any nonprohibited concomitant medications for UC during the rest of the study.

2.6 DETERMINATION OF SAMPLE SIZE

A sample size of 65 randomized subjects to each of 3 treatment groups, assuming an attrition rate of 3% per group, will allow evaluable data on approximately 63 subjects per group at Week 12. This sample size will allow for the detection of a 17-percentage point change in remission rates (assuming 5% placebo remission rate) between groups with a type I error rate of < 0.05, and 80% power. The placebo remission rate was estimated based on the results of a recent moderate-to-severe UC study by Sandborn (2016). The study found that the placebo induction clinical remission rate, as determined by a 4 component Mayo score ≤2 with no subscore >1, was 6%. Accounting for the differences in study population and scoring

method (4 component vs 3 component Mayo score), a 5% placebo rate was assumed.

2.7 METHOD OF TREATMENT ASSIGNMENT

Eligible subjects will be randomly assigned in a 1:1:1 ratio to 1 of 3 treatment groups: BT-11 low-dose (500mg), BT-11 high-dose (1,000mg), or placebo through an interactive web-based response system (IWRS). After informed consent has been documented, all subjects will receive a subject number assigned by the IWRS. Following screening procedures and confirmation of all eligibility requirements, allocation to treatment group will be performed using central randomization by means of a IWRS. Randomization will be stratified by prior exposure to biologic therapy (yes/no; exposed population limited to 30% of total sample) and current oral corticosteroid use (yes/no).

2.8 BLINDING

During the induction and maintenance period, blinding will be maintained for the sponsor, contract research organization (CRO), investigators, site personnel, and subjects. After all subjects have completed the Week 12 (or EOT) visit, the final analysis of the primary endpoint, select ranked secondary endpoints and the Week 12 AEs will be conducted by an unblinded statistician. The unblinded statistician and a sponsor representative independent of the study team will be unblinded to all Week 12 results so that analyses can be done on the indicated Week 12 endpoints. Details of treatment assignment from the data lock and final analysis of Week 12 endpoints will not be shared with any blinded personnel.

After all subjects have completed the Week 30 (or EOT) visit for the maintenance analysis (after maintenance data lock) the study will be unblinded.

2.9 STUDY POPULATION

The study will include 195 subjects with mild to moderate UC. Subjects with prior exposure to biologic therapy will be limited to 30% of the total sample. After 58 subjects with prior exposure to biologic therapy have been randomized into the induction period, recruitment will be limited to biologic-naïve subjects. A detailed list of the study inclusion/exclusion criteria can be found in Sections 4.1 and 4.2 of the protocol.

2.10 SCORING INDICES

2.10.1 Mayo Score

The 3-component modified Mayo Score consists of 3 components (stool frequency, rectal bleeding, and endoscopy findings). Each component is scored from 0-3 (as defined in Appendix 1), and summed to give a 3-component modified Mayo Score (maximum of 9 points), with higher scores representing more severe disease activity. The stool frequency subscore for efficacy analysis will be based on the average score from the most recent consecutive 3-day period (excluding date of bowel preparation for endoscopy). The rectal bleeding subscore for efficacy analysis will be based on the most severe rectal

bleeding score among the 3 scores from the most recent consecutive 3-day period (as in the stool frequency). The only difference between the modified Mayo Score and the original Mayo Score is that mild friability has been removed from the definition of mild endoscopic/sigmoidoscopic disease activity (endoscopic subscore of 1) in the modified Mayo Score. The removal of mild friability from an endoscopic subscore of 1 will be used in this analysis.

Two alternative scoring methods are also used in this study: (1) A total modified Mayo score will be based on the sum of the stool frequency, rectal bleeding, endoscopy findings, and physician's global assessment (maximum of 12 points); and (2) the partial Mayo score will be based on the sum of the stool frequency, rectal bleeding, and physician's global assessment (maximum of 9 points).

Missing values in any of the subscore components of the Mayo would result in the total score also being missing. Missing values in stool frequency and rectal bleeding subscores will only be considered missing if no consecutive 3-day period of completed values exists within the 7 days prior to study visit.

2.10.2 Geboes Score

The Geboes scoring system is a stepwise ordinal grading system for histological assessment of disease severity in UC. The scoring system progressively grades disease severity by assessing 7 histological items and grading histological change as grade 0 (structural change only), 1 (chronic cell infiltrations), 2A (lamina propria neutrophils), 2B (lamina propria eosinophils), 3 (neutrophils in the epithelium), 4 (crypt destruction), and 5 (erosion or ulceration). Each of the grades is subdivided into subgrades, based upon the severity of tissue abnormalities or the extent of cell infiltration. Subgrades are assessed from the worst area of the biopsy. Higher Geboes grades are indicative of more severe disease activity.

Due to the stepwise nature of the Geboes, if missing values occur at a lower grade, the Geboes score would not be missing so long as a higher grade is not missing and non-zero. Otherwise, the Geboes score would be considered missing.

2.10.3 Robarts Histopathology Index

The Robarts Histopathology Index (RHI) was developed from a subset of grades included in the Geboes score. These 4 items include: (1) the extent of chronic inflammatory cell infiltration (weighting factor: x1), (2) neutrophils in the lamina propria(weighting factor: x2), (3) neutrophils in the epithelium(weighting factor: x3), and (4) erosions and ulceration (weighting factor: x5). Each item is scored from 0 to 3 and multiplied by the weighting factor and summed to give the overall RHI score, with total scores ranging from 0 (no disease activity) to 33 (severe disease activity).

Missing values in any of the subscore components of the RHI would result in the index score also being missing.

2.10.4 UC-100 Index

The UC-100 is a composite disease activity index derived from Mayo Score and RHI variables. The UC-100 is calculated as (1 + [16 × Mayo stool frequency subscore] + [6 × Mayo endoscopic

subscore] + [1 × RHI score]), which ranges from 1 (no disease activity) to 100 (severe disease activity).

Missing values in any of the subscore components of the UC-100 would result in the index score also being missing.

2.10.5 Inflammatory Bowel Disease Questionnaire

The Inflammatory Bowel Disease Questionnaire (IBDQ) is a disease-specific health-related quality of life (HRQOL) instrument for patients with IBD. The IBDQ covers 4 dimensions: bowel symptoms (10 items), systemic symptoms (5 items), emotional function (12 items), and social function (5 items). Items are scored on a 7-point Likert scale and summed for a total global score in the range 32 to 224 (with higher scores indicating better HRQOL).

So long as no more than three items and no more than two items within a dimension are missing, the IBDQ will still be summed and then reweighted so that the range of possible scores are maintained. Otherwise, the IBDQ will be considered missing.

3 STATISTICAL METHODS

3.1 General Considerations

Any change to the data analysis methods described in the SAP will require an amendment only if it changes a principal feature of the SAP. Any other change to the data analysis methods described in the SAP and the justification for making the change will be described in the clinical study report. Additional exploratory analyses of the data will be conducted as deemed appropriate and noted in the clinical study report.

Descriptive statistics for continuous variables will include number of observations (N), mean, standard deviation (SD), median, minimum and maximum. Means, medians, standard deviations, and 95% confidence interval where applicable, will be presented to one more decimal place than the raw data. Minimum and maximum will be presented to the same number of decimal places as the raw data. In general, the maximum number of decimal places reported shall be four for any summary statistic.

Categorical data will be summarized in terms of the number of subjects eligible to provide data at the relevant time point (n), frequency counts and percentages. Percentages will be presented to one decimal place. Percentages will be calculated using n (the number of observations including missing values) as the denominator.

Baseline will be defined as the last non-missing assessment recorded on or prior to the date and time of first study drug dosing at Week 0 (Visit 2). Change from baseline will be calculated as the visit value of interest minus the baseline value. If all baseline values are missing for a particular variable, then the change from baseline and percent change (or percent improvement) from baseline will not be calculated.

All confidence intervals (CIs) and statistical tests will be two-sided with an alpha of 0.05 unless otherwise specified. P-values which are greater than or equal to 0.001, and less than or equal to 0.999,

will be presented to three decimal places. All other p-values which are less than 0.001 will be presented as '<0.001', while p-values greater than 0.999 will be presented as '>0.999'. CIs will be presented to one more decimal place than the raw data.

Efficacy analyses will be conducted on all randomized subjects who had at least one dose of study drug (placebo, low dose BT-11 or high dose BT-11). In keeping with the philosophy of the mITT concept, subjects will be analyzed according to the treatment to which they were assigned regardless of any errors of dosing. A robustness analysis of the primary and secondary efficacy endpoints using the Per Protocol (PP) population will be also conducted. All safety analyses will be conducted on the Safety Population. These populations are defined in Section 3.2.

Variables will be analyzed in the original scale on which they are measured.

3.1.1 Summary Tables

All summary tables will be presented by treatment group. Where applicable, data will be summarized by study visits. Summaries and analyses will be presented separately for the induction, maintenance, and open-label extension dosing periods.

3.1.2 By-Subject Listings

By-subject listings will be sorted by study site, treatment allocation, and visit. Numeric data will be listed to the same number of decimal places as recorded on the CRF or other data source. For derived variables, the number of decimal places will be listed as specified. The decimal points will be aligned in the data listings. For the tables and listings in the appendices, "Page x of X" will appear on each page.

By-subject listing will be generated for each SDTM dataset. If necessary, separate listing for derived variables (that are derived from the SDTM datasets) will be generated. Wherever applicable, listings will include the period of the study during which the event or measurement occurred.

3.2 ANALYSIS POPULATIONS

The study will include 195 subjects with mild to moderately active UC. Subjects with prior exposure to biologic therapy will be limited to 30% of the total sample. After 58 subjects with prior exposure to biologic therapy have been randomized, recruitment will be limited to biologic- naïve subjects.

Data collected on subjects who are screened but not randomized, i.e. study medication not received, will not be included in any analyses. The number of subjects screened will be reported.

3.2.1 Safety Analysis Set

All subjects who received at least 1 dose of study drug will be included in the safety population. Subjects will be analyzed according to the treatment they received regardless of indented treatment.

3.2.2 Modified Intent-To-Treat (mITT) Analysis Set

All subjects who received at least 1 dose of study drug will be included in the mITT population (also

referred to as the 'full analysis set' in the study protocol). In the event of study drug administration error, analyses on the mITT population will be performed according to the intended treatment group. The mITT population will be used for the efficacy analyses.

3.2.3 Per-Protocol Analysis Set

All mITT subjects who do not have any major deviations from protocol will be included in the Per Protocol (PP) population. Major protocol deviations will be reviewed prior to unblinding. The PP population will be used for sensitivity analyses of the primary and secondary endpoints.

3.2.4 Induction Responders Analysis Set

The Induction Responders analysis set is the subset of the mITT population that will include those subjects who have clinical response and/or clinical remission at Week 12 and receive at least 1 dose of maintenance study drug. The Induction Responders analysis set will be used for analyses during the maintenance period.

3.2.5 Subpopulations

There are several subgroups defined for this study. Details of these can be found in Sections 3.12.4 and 3.12.5.

3.3 ESTIMANDS

For the primary and key secondary endpoints, the estimand of interest is the population proportion of the respective remission/response/rate definition.

3.3.1 Target Population

The target population is characterized through the inclusion and exclusion criteria that has been outlined in the BT-11-201 protocol Section 4. Subjects may be male or female with mild to moderate UC, as defined by a total Mayo Score of 4 to 10 inclusive at baseline with a MES $\geq$ 2 (confirmed by central reader).

3.3.2 Variables

The primary endpoint is clinical remission rate at Week 12, using the 3-component modified Mayo Score as a rectal bleeding subscore of 0, a stool frequency subscore of 0 or 1, and an endoscopic subscore of 0 or 1. Followed by several key secondary endpoints, that can be found in Section 2.1.2.

3.3.3 Intercurrent Events

The following potential intercurrent events will be considered:

- If a subject fails to taper off oral corticosteroids, the while on treatment strategy will be used to analyze data collected up to and including week 12. All other endpoints that are collected after Week 12 will be treated as missing values using the hypothetical strategy.

- Subjects that initiate prohibited concomitant medications will not be removed from the analysis (treatment policy strategy).
- Subjects that are noncompliant with treatment will not be removed from the analysis (treatment policy strategy).
- If a subject discontinues from treatment, endpoints that are assessed by differences in proportions will be assessed as non-responders, using the composite strategy. Endpoints that are assessed by the difference in means will be treated as missing values through the hypothetical strategy.

The motivation behind these strategies is to quantify the treatment effect of BT-11 under the situations where:

- Any potential confounders, as defined in the inclusion and exclusion criteria, are removed since these could lead to a dilution of the treatment effect. The study sample, overall, remains representative of the subject population.
- Failure to taper off oral corticosteroids by Week 18 does not necessarily reflect the subject's lack of response to treatment; therefore, response will not be imputed.
- Ignoring initiation of prohibited concomitant medications is consistent with the mITT strategy; however, these may also be masking nonresponse.
- Ignoring small levels of treatment noncompliance is in line with the mITT strategy.
- BT-11 was taken for 12 weeks as specified; however, situations where the patient cannot complete the course of treatment cannot be ignored.

3.3.4 Summary Measure

The summary measure for the primary endpoint and the categorial secondary endpoints will be the difference in proportions between 500 mg and 1,000 mg treatment groups compared to the placebo group using the Cochran-Mantel-Haenszel test, stratified by previous use of biologic therapy (yes/no) and oral corticosteroid use at baseline (yes/no). The summary measure for the continuous secondary endpoints will be the difference in means between 500 mg and 1,000 mg treatment groups compared to the placebo using a mixed model repeated measure analysis method. The model will include fixed effects for treatment, visit, previous biologic use, previous corticosteroid use, and treatment-by-visit interaction.

3.4 ADJUSTMENTS FOR COVARIATES

The randomization at the beginning of the Induction Period is stratified by prior exposure to biologic therapy for UC (yes/no) and corticosteroid use at baseline (yes/no). Unless otherwise specified and where applicable, all efficacy analyses will account for the stratification factors through three strata. These three strata are biologic naïve with no corticosteroid use at baseline, biologic naïve with corticosteroid use at baseline and biologic exposed.

Unless otherwise specified, analyses will be performed utilizing the methodology and covariates described in Section 3.12.

3.5 HANDLING OF MISSING DATA

All available efficacy and safety data collected for the study will be included in data listings and/or summary tables. No imputation of values for missing safety data will be performed. Missing values in dichotomous efficacy endpoints will be imputed as non-responses. Missing values in continuous efficacy endpoints will be handled implicitly in the mixed model regression analysis.

As a sensitivity analysis, continuous efficacy endpoints will also be analyzed with missing values imputed using the Last Observation Carried Forward (LOCF) approach.

3.5.1 Non-Responder Imputation (NRI)

The primary outcome is the proportion of subjects with clinical remission at 12 weeks. For this and other categorical efficacy endpoints (for example, clinical remission (week 30), clinical response (week 12, 30), and endoscopic remission (week 12, 30)), non-responder imputation (NRI) will be used for missing clinical assessment values. Specifically, all subjects who discontinue from the study at any time prior to the assessment timepoint (regardless of reason) or fail to have an adequate efficacy assessment at that timepoint will be considered a non-responder.

The NRI may be applied at any time point specified for analysis.

3.5.2 Last-Observation Carried Forward (LOCF)

An LOCF analysis will be performed on continuous efficacy analysis, such as Partial Mayo Clinic Score at Weeks 2, 6, and 12.

This is both a sensitivity analysis, and an analysis for regulatory agencies that prefer this approach. The last non-missing post-baseline observation before discontinuation will be carried forward to the corresponding endpoint for evaluation. Randomized subjects without at least one post-baseline observation will not be included for evaluation.

3.6 MULTIPLE COMPARISONS/MULTIPLICITY

A closed hierarchical procedure will be used to control for multiple comparisons. The order of testing will be high-dose versus placebo for the primary endpoint at Week 12 as the first test. If this result is significant at the 2-sided P <0.05 then the low-dose versus placebo at Week 12 will be tested, followed by subsequent ranked key secondary endpoints (with ranking specified in Section 3.12.2. In this regard, the first ranked secondary endpoint is similarly tested first for the high dose and subsequently for the low dose if p<.05 for the high dose for the first ranked secondary endpoint. Testing continues in a similar manner for the subsequent ranked secondary endpoints with high dose tested prior to low dose. If at any point in this sequential procedure the P <0.05 is not met, the testing procedure will be terminated. All

subsequent analyses would be considered exploratory.

3.7 SUBJECT DISPOSITION

All subjects who discontinue from the study will be identified, and the extent of their participation in the study will be reported. If known, a reason for their discontinuation will be given.

Total screened subjects and screen failed subjects will be expressed as overall counts. Counts and percentages of subjects randomized, discontinued, and completed will be summarized by treatment group. The reasons for study discontinuation will be summarized by treatment group.

Summaries of subject disposition will be based on all subjects screened in the study. The number of subjects in each of the analysis populations will be displayed. The reasons for study discontinuation will be summarized. A listing of subjects excluded from any population, if any, will be provided along with the reason for the exclusion.

3.8 SUBJECT CHARACTERISTICS

The subject's year of birth, gender, weight, height, previous biologic treatment, previous corticosteroid therapy treatment, and other demographic characteristics are collected at the screening visit. Age and body mass index will be calculated.

Only the year of birth is collected at screening. For the purpose of age calculation, the month and day of birth will be imputed as July 01, of the year of birth. Age is the computed as follows:

$$\text{AGE} = (\text{Informed Consent Date} - \text{Date of Birth})/365.25$$

Demographic and baseline characteristics (including age, gender, race and ethnicity) will be summarized for each treatment group.

Certain characteristics, that are collected at baseline or after baseline but not summarized in the demographic summary, will be reported as a listing.

No inferential analysis for the comparability of baseline covariates across treatment groups will be performed.

3.9 MEDICAL, SURGICAL, AND MEDICATION HISTORY

Medical, surgical and medication history will be collected retrospectively at the screening visit. Medical and surgical history will include UC-related complications, other significant conditions or diseases relevant to the disease under study, previous surgeries for UC, and the extent and duration of disease at baseline.

Medication history will include medications relevant to the eligibility criteria and any others that were stopped at or within 3 months prior to signing of the informed consent. All prior biologic medication history for the treatment of UC (with duration of previous biologic use, name of therapy, reason for discontinuation) will be collected.

Medical, surgical and medication history will be presented in by-subject listing which includes Subject ID, treatment group, medical, surgical and medication history.

3.10 TREATMENT COMPLIANCE

Subjects will be randomized to receive BT-11 low-dose (500 mg), BT-11 high-dose (1,000 mg), or placebo once-daily for 12 weeks. Compliance with study drug will be monitored by site personnel. The first dose will be given in the clinic. If a subject is continually noncompliant with study drug, the investigator may deem it appropriate to withdraw the subject from the study. Subjects will be reminded about dosing requirements at study visits.

Study treatment administration and compliance will be listed for all entered subjects. Number and percentage of subjects who missed doses since last visit will be summarized by treatment group for each treatment period.

No subject will be excluded from the mITT population as a consequence of significant noncompliance.

3.11 CONCOMITANT MEDICATIONS

Concomitant medications will be recorded at each visit.

For the purposes of subgroup identification, medications will be classified into anatomical therapeutic chemical (ATC) drug classes using the latest version of the World Health Organization (WHO) drug dictionary. Summaries of Concomitant and Prior medication will utilize WHO drug preferred names.

Medication start and stop dates will be compared to the date of first dose of treatment in each treatment period to allow medications to be classified as Concomitant for each treatment period. Concomitant medication use that starts and stops during the baseline/screening period will be included in summaries of the induction period.

Prior medications are those medications that start and stop prior to the date Informed Consent (IC). Concomitant medications are those medications that start after the IC date. Concomitant medications are assigned to the treatment period in which they are actually ongoing. For example, if a subject is receiving concomitant medication during the Induction Dosing Period but has a stop date during the Induction Dosing Period, the same medication would not be listed as a concomitant medication during the latter periods unless subject has a new start date.

3.12 EFFICACY ANALYSES

Efficacy analyses will primarily be based on the mITT population. Efficacy analyses will also be performed on the PP analysis set for robustness purposes. Statistical tests will be 2-sided and performed at the 0.05 level of significance.

3.12.1 Primary Efficacy Analysis

The primary efficacy parameter, clinical remission rate at Week 12, as defined as a rectal bleeding

subscore of 0, a stool frequency subscore of 0 or 1 and an endoscopic appearance subscore of 0 or 1, will be assessed using the mITT population.

The proportion of subjects with clinical remission at Week 12 will be compared between each BT-11 treatment group and placebo using a Cochran-Mantel-Haenszel test; two binary stratification factors are used to give four strata: (1) biologic naïve (yes/no) and (2) corticosteroid use at baseline (yes/no). The stratified relative risk versus placebo of each BT-11 group will be reported along with 95% Wald confidence limits. An associated p<.05will be used to define statistical significance.

Missing values in the primary endpoint will be imputed as non-responders. Analysis will also be performed on the PP analysis set as a sensitivity analysis.

3.12.2 Secondary Efficacy Analysis

Dichotomous secondary endpoints will be analyzed as described in the primary outcome measure section (Section 3.12.1).

Continuous secondary endpoints will be summarized by treatment group using descriptive statistics (mean, median, standard deviation, minimum, and maximum) of the values at each visit and the change from baseline to each post-baseline visit.

The main analysis methods for continuous secondary endpoints at two or more post-baseline visits will be based on the mixed model repeated measures analysis method. The model will include fixed effects for treatment, visit, previous biologic use, previous corticosteroid use, study region (Europe, USA), baseline value, and treatment-by-visit interaction. This analysis will be conducted by maximum likelihood methods using the SAS PROC MIXED procedure. While measurements taken at different time points are expected to correlate, we also assume that measurements taken at adjacent time points are more correlated than measurements further apart, and therefore we will use an unstructured covariance structure. If the model with unstructured covariance does not converge or it is determined to be inappropriate as outlined in Guerin and Stroup (2000), then other covariance structures such as a heterogeneous Toeplitz, AR(1), and compound symmetry will be considered to model the within-subject errors. The model with smaller Akaike's information criterion will be chosen for the principal analysis. Denominator degrees of freedom will be estimated using the Kenward-Roger approach.

For continuous secondary endpoints at only one post-baseline visit, analysis of covariance (ANCOVA) will be applied with the model including treatment, previous biologic use, previous corticosteroid use, study region (Europe, USA), and baseline value.

All collected efficacy data will also be presented in by-subject listings.

Key secondary efficacy assessments (ranked for hierarchical testing procedure as described in Section 3.6) are:

- Endoscopic remission rate at Week 12 (MES of 0 or 1)
- Endoscopic response rate at Week 12 (defined as a decrease from baseline in MES of at least 1 point)

- Clinical response rate at Week 12 (defined as decrease from baseline in Mayo score of at least 3 points and at least 30 percent, with an accompanying decrease in the subscore for rectal bleeding of at least 1 point or an absolute subscore for rectal bleeding of 0 or 1)
- Mucosal healing rate at Week 12 as defined by a MES of 0 or 1 and a Geboes Histologic Index score of less than 3.1
- Histologic remission rate at Week 12 as defined by a Geboes Histologic Index score of less than 3.1

The same data handling rules that were specified for the primary endpoint analysis will also be used for each of these key secondary endpoints. In particular, missing values in the key secondary endpoint will be imputed as non-responders.

Other secondary efficacy assessments are:

- Mean change in 3-component modified Mayo score from baseline to Week 12
- Mean change in Partial Mayo Clinic Score from baseline to Weeks 2, 6, and 12
- Mean change in Mayo rectal bleeding subscore from baseline to Weeks 2, 6, and 12
- Mean change in Mayo stool frequency subscore from baseline to Weeks 2, 6, and 12
- Mean change in MES from baseline to Week 12
- Mean change in RHI scores from baseline to Week 12
- Mean change in fecal calprotectin from baseline to Weeks 2, 6, and 12
- Clinical remission rate at Week 12, based on alternate definition ("MCS clinical remission") of total Mayo Score ≤ 2 with all subscores ≤ 1
- Normalization of fecal calprotectin at Weeks 2, 6, and 12 in subjects with abnormal fecal calprotectin at baseline (abnormal defined as fecal calprotectin> 250 mg/kg)
- Mean change in hs-CRP from baseline to Weeks 2, 6, and 12
- Normalization of hs-CRP at Weeks 2, 6, and 12 in subjects with abnormal hs-CRP at baseline (abnormal hs-CRP defined as hs-CRP > 3.0 mg/L)
- Mean change in UC-100 score from baseline to Week 12
- Mean change in Robarts PRO from baseline to Week 12
- Mean change in IBDQ score from baseline to Week 12
- BT-11 PK parameters (concentration in serum, feces, and tissue)

3.12.3 Exploratory Efficacy Analysis

The exploratory objectives of this study are to evaluate the following objectives through Week 30 to the

extent feasible. Categorical exploratory endpoints will be analyzed as described in Section 3.12.1 to the extent to which such analyses are applicable to the corresponding data structure. Continuous exploratory endpoints will be analyzed as described in Section 3.12.2 to the extent to which such analyses are applicable to the corresponding data structure. In this regard, patients who are not clinical responders at Week 12 do not have any assessments subsequent to Week 12, with this implying that only statistical descriptions are possible for some of the subsequently shown exploratory endpoints.

To evaluate the effects of BT-11 on disease activity as measured by symptoms, endoscopy, histology, and biomarkers during up to 30 weeks of maintenance therapy:

- Durable clinical remission (defined as clinical remission at both Week 12 and Week 30)
- Durable clinical remission (defined as MCS clinical remission at both Week 12 and Week 30)
- Durable clinical response (defined as clinical response at both Week 12 and Week 30)
- Endoscopic remission rate at Week 30 (MES of 0 or 1)
- Endoscopic response rate at Week 30 (defined as a decrease from Week 12 in MES of at least 1 point)
- Corticosteroid-free clinical remission at Week 30
- Corticosteroid-free endoscopic remission at Week 30
- Mucosal healing rate at Week 30 as defined by a MES of 0 or 1 and a Geboes Histologic Index score of less than 3.1
- Histologic remission rate at Week 30 as defined by a Geboes Histologic Index score of less than 3.1
- Clinical response rate at Week 30 (defined as decrease from Week 12 in Mayo score of at least 3 points and at least 30 percent, with an accompanying decrease in the subscore for rectal bleeding of at least 1 point or an absolute subscore for rectal bleeding of 0 or 1)
- Mean change in 3-component modified Mayo score from baseline to Week 30.
- Mean change in Partial Mayo Clinic Score from baseline to Weeks 18, 24, and 30
- Mean change in Mayo rectal bleeding subscore from baseline to Weeks 18, 24, and 30
- Mean change in Mayo stool frequency subscore from baseline to Weeks 18, 24, and 30
- Mean change in MES from baseline to Week 30
- Mean change in RHI scores from baseline to Week 30
- Mean change in fecal calprotectin from baseline to Weeks 18, 24, and 30
- Normalization of fecal calprotectin at Weeks 18, 24, and 30 in subjects with abnormal fecal calprotectin at baseline (abnormal defined as fecal calprotectin > 250 mg/kg)

- Mean change in high-sensitivity C-reactive protein (hs-CRP) from baseline to Weeks 18, 24, and 30
- Normalization of hs-CRP at Weeks 18, 24, and 30 in subjects with abnormal hs-CRP at baseline (abnormal defined as hs-CRP > 3.0 mg/L)
- Mean change in UC-100 score from baseline to Week 30
- Mean change in Robarts PRO from baseline to Week 30

To assess health-related quality of life during up to 30 weeks of maintenance therapy:

- Mean change in IBDQ score from baseline to Week 30, comparing BT-11 to placebo

To assess the PK parameters of BT-11 during up to 30 weeks of maintenance therapy:

- In the active BT-11 treatment group: BT-11 concentration in serum, feces, and tissue

To evaluate safety during up to 30 weeks of maintenance therapy:

- Frequency and severity of adverse events compared to placebo
- Changes in clinical chemistry and hematology from Week 12
- Results of vital signs and physical examination
- ECG findings

To evaluate target engagement and mechanism of action: (Details of this analysis will be presented in a separate document)

- Immunohistochemistry on formalin-fixed paraffin-embedded colonic mucosal tissue biopsies
    - MPO (marker for neutrophil infiltration in tissue)
    - CD25 (regulatory T cell marker)
    - FOXP3 (regulatory T cell marker)
    - LANCL2 (target)
- Multiplex cytokine analysis in serum and tissue
    - TNFα, IFNγ, IL-4, MCP1, MIP1a, IL-8, IL-6 (pro-inflammatory cytokines)
    - IL-10 (anti-inflammatory cytokine)
- Transcriptomic analysis on colonic mucosal tissue biopsies using quantitative polymerase chain reaction (qPCR)
    - TNFα, IFNγ, IL-4, MCP1, IL-8, IL-6 (pro-inflammatory cytokines)
    - IL-10, FOXP3 (anti-inflammatory cytokines)
    - LANCL2 (target)

- Flow cytometry on colonic mucosal tissue biopsies
  - Th1 cells
  - Neutrophils
  - IL-10-producing cellular subsets including CX3CR1+ macrophages
  - Treg cells

To explore association of drug exposure in colonic mucosal tissue biopsies with clinical, endoscopic, histopathologic, cellular, and molecular outcomes: (Details of this analysis will be presented in a separate document)

- Exposure-response analysis

3.12.4 Subgroup Analysis

Key primary and secondary endpoints may be subgrouped for exploratory purposes. These subgroups could include age (<55 vs ≥55), gender, race (Caucasian vs non-Caucasian), and subject region (North America vs Europe). Categorical exploratory endpoints will be analyzed as described in Section 3.12.1. Continuous endpoints for subgroups will be analyzed as described in Section 3.12.2.

3.12.5 Sensitivity Analysis

Three sensitivity analyses may be applied to the study data.

If there is at least one subject who initiated the use of prohibited concomitant medications during the course of the trial, the primary and key secondary endpoints will be analyzed with an additional study population that has the prohibited concomitant medication taking subjects treated as non-responders (for categorical endpoints) or as missing (for continuous endpoints).

To assess the sensitivity of the study conclusions to the subject's baseline disease severity, the primary and key secondary endpoints will be assessed with an additional study population in which no baseline Mayo subscore satisfies the criteria for clinical remission; that is, subjects with a baseline where stool frequency ≥2, rectal bleeding ≥1, and endoscopic evaluation ≥2. To further evaluate the impact of baseline disease severity, subjects will be subgrouped by baseline 4-component Mayo score; specifically, the primary and key secondary endpoints will be evaluated separately for those subjects with a baseline Mayo score in the lower half of the study range (i.e. below the median value) and those subjects with a baseline Mayo score in the upper half of the study range (i.e. at or above the median value).

3.13 SAFETY ANALYSES

Safety analyses will be based on the SAF population. Assessments will be made by evaluating all reported AEs, physical examination findings, and changes in laboratory analytes, ECGs, and vital signs (including body weight). Safety data will be presented by treatment groups. Descriptive statistics will be presented. No formal statistical tests will be performed.

3.13.1 Extent of Exposure

A by-subject listing of exposure will be provided. No inferential analysis for comparison between treatment arms will be performed.

3.13.2 Adverse Events

All AEs will be classified according to the Medical Dictionary for Regulatory Activities (MedDRA) dictionary. AEs will be summarized by the number and percentage of subjects experiencing any AEs, any treatment-related AEs, any SAEs, any treatment-related SAEs, and any AEs leading to study drug interruption/discontinuation. Adverse events will also be summarized by 1) system organ class and preferred term; 2) system organ class, preferred term, and severity; and 3) system organ class, preferred term, and relationship to study drug. If a subject experienced more than 1 AE for a given preferred term, severity will be defined as the severity of the most severe event and relationship to study drug is the relationship of the most related event.

To compare between treatment groups the relative risk and corresponding 95% confidence interval of each BT-11 group vs placebo will be calculated for incidence of:

- AEs that lead to study drug discontinuation
- SAEs
- Severe AEs
- AEs determined to be possibly or more likely related to study treatment

Given the well tolerated benign safety profile of BT-11 in Phase 1 and animal toxicity studies, no AEs of special interest exist in regard to mechanism of action and previous clinical and nonclinical experience.

3.13.3 Physical Examination

Physical examination is performed at the screening visit. All abnormal findings will be presented in a by-subject listing.

3.13.4 Vital Signs

Vital sign measurements will be obtained at each clinic visit and will include weight, temperature, blood pressure readings, and pulse. Body mass index (BMI) will be calculated from height and weight measurements.

For analyses, values at each visit and change from baseline to each visit will be summarized by treatment group. Baseline will be the last non-missing observation in the screening period. The following summary statistics will be included in the table: N, mean, standard deviation, minimum, and maximum.

Vital signs data will also be presented in a by-subject listing, which will include subject identifier, treatment group, vital sign collection date/time, test name, and test result.

3.13.5 Clinical Laboratory Tests

For analyses, values at each visit and change from baseline to each visit will be summarized by treatment group. Baseline will be the last non-missing observation in the screening period (induction) or the last non-missing observation prior to enrolment (maintenance). The following summary statistics will be included in the table will include N, mean, standard deviation, minimum, and maximum.

In addition to the central laboratory safety tests, a serum pregnancy test will be carried out at the screening visit and urine pregnancy tests at other visits as applicable in female subjects of childbearing potential.

In the by subject listings, laboratory values which fall outside the reference ranges and abnormal findings will be flagged. The listing will include subject identifier, treatment group, laboratory collection date/time, analyte name, and analyte finding.

3.13.6 Electrocardiograms

ECG measurements are obtained at the screening visit, Visit 5 (Week 12)/ induction EOS and Visit 9 (Week 30)/ maintenance EOS. Single 12-lead ECG will be obtained using an ECG machine that automatically reports heart rate, PR, QRS, QT, and Fridericia's corrected QT (QTcF) intervals. For analyses at the induction period, values at each visit and change from baseline to each visit will be summarized in tables. For analyses at the maintenance period, values at Visit 9 and change from Visit 5 will be summarized.

ECG measurements will also be presented in the by-subject listings, which includes subject identifier, treatment group, ECG date/time, test name, and test result.

If any clinically significant ECG measurement occurs, it will be recorded as an AE. ECG data will not be analyzed in ECG tables or listings.

3.13.7 Liver Safety Monitoring

The number and percentages of subjects with the following elevations in hepatic laboratory tests at each scheduled laboratory visit will be summarized between treatment groups. Separate summary table will be provided for both induction and maintenance periods.

- ALT ≥ 3x upper limit of normal (ULN)
- AST ≥ 3x upper limit of normal (ULN)
- Total bilirubin level (TBL) ≥ 2xULN
- International normalized ratio (INR) > 1.5

Subjects with the above elevations will be presented in the by-subject listings.

3.14 PROTOCOL VIOLATIONS

Review of all major and minor protocol violations will be performed on an ongoing basis during the

conduct of the study. All protocol exemptions/violations identified will be tracked. All protocol exemptions and violations that require medical review will be listed.

3.15 NUMBER AND TIMING OF ANALYSES

Final analysis for the primary and select key secondary endpoints will be conducted after all subjects have reached Week 12. As this will be the final analysis for these endpoints, no adjustment for type I errors will be made. Eligible subjects who meet the definition of clinical response and/or clinical remission at Week 12 will be classified as responders and will continue on to the maintenance period. Subjects who are non-responders at Week 12 are eligible for a separate open-label extension (OLE) study, if available.

After all subjects have completed the Week 12 (or EOT) visit for the induction phase analysis lock, select members of the analysis team and the sponsor will be unblinded to the subject treatment groups so that analysis can be done on the corresponding induction phase data structure through Week 12. Details of treatment assignment from the final analysis for the induction phase data structure will not be shared with any blinded personnel until the study is complete (including the maintenance period of the trial).

Final analysis of the induction and maintenance phase data structure will be conducted after all subjects who entered the maintenance period have completed the Week 30 (or EOT) visit.

4 REPORTING OF DEVIATIONS TO ORIGINAL SAP

All deviations from the original SAP will be reported in the clinical study report.

5 REVISION HISTORY

Version	Effective Date	Reason
V01	03NOV2020	New

APPENDIX 1 MODIFIED MAYO SCORE

Component	Mayo Score
Stool frequency	
Normal number of stools for this subject	0
1 to 2 stools more than normal	1
3 to 4 stools more than normal	2
5 or more stools than normal	3
Subscore, 0 to 3	
Rectal bleeding	
No blood seen	0
Streaks of blood with stool less than half the time	1
Obvious blood (more than just streaks) or streaks of blood with stool most of the time	2
Blood alone passes	3
Subscore, 0 to 3	
Findings on endoscopy	
Normal or inactive disease	0
Mild disease (erythema, decreased vascular pattern, **no friability**)	1
Moderate disease (marked erythema, absent vascular pattern, friability, erosions)	2
Severe (spontaneous bleeding, ulceration)	3
Subscore, 0 to 3	
Physician's global assessment	
Normal	0
Mild disease	1
Moderate disease	2
Severe disease	3
Subscore, 0 to 3	
Total score	**0-12**

Source: Schroeder KW, Tremaine WJ, Ilstrup DM. Coated oral 5-aminosalicylic acid therapy for mildly to moderately active ulcerative colitis. A randomized study. N Engl J Med. 1987;317(26):1625-1629.